Epidemics 14 (2016) 71–82

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Expanding vaccine efficacy estimation with dynamic models fitted to cross-sectional prevalence data post-licensure

Erida Gjini[a,*], M. Gabriela M. Gomes[b,c,d]

[a] Instituto Gulbenkian de Ciência, Apartado 14, 2781-901 Oeiras, Portugal
[b] CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade de Porto, Portugal
[c] Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil
[d] Liverpool School of Tropical Medicine, Liverpool, United Kingdom

ARTICLE INFO

Article history:
Received 8 May 2015
Received in revised form 2 November 2015
Accepted 25 November 2015
Available online 9 December 2015

Keywords:
Vaccination model
Strain replacement
Co-infection
Competition
ODE parameter inference

ABSTRACT

The efficacy of vaccines is typically estimated prior to implementation, on the basis of randomized controlled trials. This does not preclude, however, subsequent assessment post-licensure, while mass-immunization and nonlinear transmission feedbacks are in place. In this paper we show how cross-sectional prevalence data post-vaccination can be interpreted in terms of pathogen transmission processes and vaccine parameters, using a dynamic epidemiological model. We advocate the use of such frameworks for model-based vaccine evaluation in the field, fitting trajectories of cross-sectional prevalence of pathogen strains before and after intervention. Using SI and SIS models, we illustrate how prevalence ratios in vaccinated and non-vaccinated hosts depend on true vaccine efficacy, the absolute and relative strength of competition between target and non-target strains, the time post follow-up, and transmission intensity. We argue that a mechanistic approach should be added to vaccine efficacy estimation against multi-type pathogens, because it naturally accounts for inter-strain competition and indirect effects, leading to a robust measure of individual protection per contact. Our study calls for systematic attention to epidemiological feedbacks when interpreting population level impact. At a broader level, our parameter estimation procedure provides a promising proof of principle for a generalizable framework to infer vaccine efficacy post-licensure.

1. Introduction

Mathematical epidemiological models for the dynamics of microparasite infections have a long history of development and use in the design and optimization of intervention programmes (Anderson et al., 1992). Yet, many challenges remain in applying such models retrospectively to interpret and quantify intervention effects in host–pathogen systems (Keeling, 2005; O'Hagan et al., 2014; Wikramaratna et al., 2014; Goeyvaerts et al., 2015). It is of public interest to quantify the relative effectiveness of different control strategies, assess the ongoing changes in transmission dynamics following such interventions, and optimize their design through a cost–benefit analysis for the future. In this paper, our focus is on vaccination as a transmission-reducing intervention, and more specifically, in the context of endemic pathogens. Although the amount of data available from epidemiological trials, cross-sectional and longitudinal surveys is vast and rapidly increasing, our understanding and interpretation of such data on the basis of transmission mechanisms and epidemiological feedbacks is limited. This is apparent for many pathogen systems, including *Streptococcus pneumoniae* bacteria, human papillomaviruses, dengue, malaria, influenza and rotaviruses. Currently several vaccines are being used or contemplated to control these pathogens around the world (Comanducci et al., 2002; Insinga et al., 2007; del Angel and Reyes-del Valle, 2013; Sabchareon et al., 2012; Agnandji et al., 2011; Black et al., 2000), and assessing their efficacy is crucial.

Conceptual models can play a key role in this assessment, first by clearly defining the measures of interest, secondly, by distinguishing individual from population indicators, and thirdly, by enabling us to anticipate future outcomes of vaccination programmes. An important vaccine parameter is efficacy against pathogen acquisition, defined as reduction in the probability of infection per contact of each vaccinated individual (Haber et al., 1991). Before a vaccine is introduced, vaccine efficacy estimation is typically performed through randomized controlled trials, involving a subset of a given

* Corresponding author.
E-mail address: egjini@igc.gulbenkian.pt (E. Gjini).

http://dx.doi.org/10.1016/j.epidem.2015.11.001

population (Halloran et al., 2010). Such vaccine evaluation studies use 1 − RR (1 minus risk ratio), as a measure of efficacy, where RR is some estimate of relative risk in vaccinated vs. non-vaccinated individuals. This tends to ignore indirect effects (Halloran et al., 1991), such as the changes in transmission mediated by the intervention, which while in the time and coverage of trials are indeed expected to be negligible, are not quite negligible when mass-immunization is in place (Shim and Galvani, 2012).

The assessment of vaccines post-licensure is also of interest, and here is where dynamic mathematical models can be useful, alongside statistical approaches (Biondi and Weiss, 2015; Crowe et al., 2014; Andrews et al., 2014). There are several reasons for why such a-posteriori assessment is important. First, only a dynamic model can properly link pre-licensure vaccine expectations and observed outcomes in a population undergoing immunization, thereby providing a validity test for the numerical estimates of *vaccine efficacy* obtained from trials, and a validity test for the public-health projections made a priori regarding *effectiveness*, or population level impact. Second, only a dynamic model can take into account in a mechanistic manner the time since the onset of the vaccination programme, regardless of equilibrium requirements (Rinta-Kokko et al., 2009), and consider the actual vaccine coverage in a given setting. Third, in the context of multi-strain pathogens, where multivalent vaccines target a subset of pathogen types, only a dynamic model can properly implement the nonlinear interactions between pathogen types (Lipsitch, 1997; Martcheva et al., 2008), arising through direct competition, cross-immunity or asymmetric vaccine protection.

Although there has been recognition of the importance of dynamic transmission models for vaccine assessment (Shim and Galvani, 2012), few studies so far have attempted to infer vaccine efficacy fitting dynamic models to temporal prevalence trajectories post-vaccination (Choi et al., 2011; Gjini et al., 2016). Other approaches have suggested that prevalence odds ratios may be more suitable than prevalence ratios to determine vaccine efficacy, and that special attention must be given to the time of sampling post-vaccination (Scott et al., 2014). Another study by Omori et al. (2012) has used dynamic models (SIS and SIR) to illustrate the bias in odds-ratio estimators of vaccine efficacy for two competing pathogen types, but their estimation was based on prevalences at endemic steady state only, posing a strong restriction on the method. A recent study by van Boven et al. (2013) deals with vaccine efficacy estimation in an epidemic scenario, and applies a dynamic modelling framework to mumps outbreak data in the Netherlands.

Here, we advocate a similar dynamic spirit in the context of endemic diseases. We propose a novel approach to vaccine efficacy estimation using cross-sectional prevalence data integrated within dynamic mathematical models. This enables a deeper understanding of vaccine performance in the field, as mediated by transmission intensity, competition between pathogen subtypes and host factors. When vaccine coverage is high, the transmission cycle encompasses vaccinated and non-vaccinated individuals interacting through contact, thus affecting and being subject to a dynamic force of infection. With a gradually diminishing exposure to vaccine types, in polymorphic systems, subtype relative frequencies can change in the population from the combined effects of vaccination and interactions between target and non-target pathogen types. If a vaccine induces a replacement phenomenon, as it has been argued for pneumococcus (Weinberger et al., 2011) and HPV (Biondi and Weiss, 2015), vaccine efficacy against targeted pathogen strains, can be estimated while these strains are still in circulation, namely while type replacement is not yet complete, and sufficient information can be extracted. It is precisely in this intermediate dynamic phase that most vaccine observational studies are conducted, and where epidemiological feedbacks, including changes in exposure and interaction between multiple strains, are most likely to play a role.

To correctly capture all these processes, more refined mathematical frameworks are needed. This requires going beyond direct statistical comparisons, based on static data, e.g. snapshot prevalence odds ratios from observational studies (Thompson et al., 1998), or the indirect cohort method for case–control data (Andrews et al., 2011), which neglects pathogen subtype interference altogether. Even more importantly, the cohort method fails to acknowledge that the probability of infection of an individual depends on the infection prevalence in the population, i.e. on the infection status of others.

With a dynamic modelling approach, instead, the problem of constant hazard ratios (Hernán, 2010) can be circumvented, as can limitations of the indirect cohort method (Moberley and Andrews, 2014) for purposes of vaccine efficacy estimation. Furthermore, data can be interpreted relaxing the stationarity requirement and accounting for pathogen type replacement. Other statistical estimation methods such as incidence density sampling (Richardson, 2004), might also not require the assumption of stationarity, but they do not deal with competition in multi-strain pathogen systems.

The definition of vaccine efficacy that we consider in this paper has a clear biological meaning: reduction of the probability of pathogen acquisition per contact, which enables extrapolation beyond a single study population. This contrasts classical estimates of vaccine efficacy that are based on comparing attack rates in vaccinated and unvaccinated individuals (the cohort method), or those that use the vaccination status of the infected individuals relative to the population vaccination coverage (the screening method). Such vaccine efficacy indicators lack a clear biological meaning, which makes interpretation problematic, and prevents anticipation of the critical vaccination coverages needed to reach certain desired outcomes.

In this study, we argue that temporal effects of vaccination programmes can be addressed through dynamic mathematical models, where parameters of efficacy are explicitly defined in terms of underlying transmission mechanisms, and where epidemiological feedbacks among immunized and non-immunized individuals, and between pathogen strains are correctly accounted for. In the interest of simplicity and clarity, we only consider minimal epidemiological models to illustrate vaccine effects on single and multiple infection with different pathogen types, but the uncovered trends should apply in similar vein to more complex vaccination scenarios (Halloran et al., 1991, 2010). We delineate a proof-of-concept inference procedure, based on ODE model fitting, to cross-sectional data collected over different time points after vaccine implementation.

2. Materials and methods

To build intuition in our reader, initially we present susceptible-infected (SI) model frameworks accounting for one and two pathogen types, while the susceptible-infected-recovered (SIR) analogues are elaborated in the Supplementary Text S2. Then we proceed to susceptible-infected-susceptible (SIS) models with many-type pathogens, grouped according to whether they are targeted by a polyvalent vaccine or not. We always assume that the vaccine is effective against type 1 pathogen (SI/SIR models), or against pathogen subtypes in group 1 (SIS setting). The mode of action of the vaccine we consider is leaky (Halloran et al., 1991), and the vaccine efficacy is defined as the reduction in probability of infection/pathogen acquisition per contact. Notice that in this paper, we will use the terms ‘infection’ and ‘carriage’ interchangeably. As the source of prevalence data, we consider active

surveillance programmes pre- and post-vaccination in a population, whereby the carriage status and pathogen type(s) of each screened individual are determined. The basic structure of the models in the absence of an intervention is given in Fig. 1.

2.1. SI model – 1 pathogen type (n = 1)

The first model we consider for illustration is a simple susceptible-infected model with one pathogen type that is directly-transmitted (SI-1). With a continuous vaccination programme in place, the proportions of hosts in different compartments are given by:

$$\begin{cases} \text{Non-vaccinated hosts} \\ \frac{dS^0}{dt} = \mu(1-\rho) - \beta S^0(I^0+I^1) - \mu S^0 \\ \frac{dI^0}{dt} = \beta S^0(I^0+I^1) - \mu I^0 \end{cases} \quad \begin{cases} \text{Vaccinated hosts} \\ \frac{dS^1}{dt} = \mu\rho - \beta w S^1(I^0+I^1) - \mu S^1 \\ \frac{dI^1}{dt} = \beta w S^1(I^0+I^1) - \mu I^1 \end{cases}$$

where subscripts 0 and 1 indicate non-vaccinated and vaccinated host status, respectively. The parameter β is the per-capita transmission coefficient and μ the birth rate (equal to the death rate). Hosts are born with a life expectancy of $1/\mu$. The vaccination coverage at birth is ρ and vaccine efficacy is given by $1-w$, assuming a homogeneous effect (leaky vaccine). In the absence of a vaccine, for such pathogen to persist, the basic reproduction number (Heesterbeek, 2000) $R_0 = \beta/\mu$ must exceed 1, and the higher R_0 is, the higher the pathogen prevalence.

2.2. SI model with 2 competing pathogen types (n = 2)

Extending the above model to two pathogen types (SI-2), we have:

$$\begin{cases} \text{Non-vaccinated hosts} \\ \frac{dS^0}{dt} = \mu(1-\rho) - (\lambda_1+\lambda_2)S^0 - \mu S^0 \\ \frac{dI_1^0}{dt} = \lambda_1 S^0 - I_1^0\sigma_1\lambda_2 - \mu I_1^0 \\ \frac{dI_2^0}{dt} = \lambda_2 S^0 - I_2^0\sigma_2\lambda_1 - \mu I_2^0 \\ \frac{dI_{12}^0}{dt} = \sigma_1\lambda_2 I_1^0 + \sigma_2\lambda_1 I_2^0 - \mu I_{12}^0 \end{cases} \quad \begin{cases} \text{Vaccinated hosts} \\ \frac{dS^1}{dt} = \mu\rho - (w\lambda_1+\lambda_2)S^1 - \mu S^1 \\ \frac{dI_1^1}{dt} = w\lambda_1 S^1 - I_1^1\sigma_1\lambda_2 - \mu I_1^1 \\ \frac{dI_2^1}{dt} = \lambda_2 S^1 - I_2^1\sigma_2 w\lambda_1 - \mu I_2^1 \\ \frac{dI_{12}^1}{dt} = \sigma_1\lambda_2 I_1^1 + \sigma_2 w\lambda_1 I_2^1 - \mu I_{12}^1, \end{cases}$$

where $\lambda_1 = \beta(I_1^0 + I_{12}^0/2 + I_1^1 + I_{12}^1/2)$ and $\lambda_2 = \beta(I_2^0 + I_{12}^0/2 + I_2^1 + I_{12}^1/2)$. The above equations track the proportions of non-vaccinated and vaccinated hosts in 4 classes: susceptibles, S, hosts carrying pathogen type 1, I_1, hosts carrying pathogen type 2, I_2, and co-infected hosts I_{12}, where $S^0 + \sum I^0 = 1-\rho$ and $S^1 + \sum I^1 = \rho$ respectively for non-vaccinated and vaccinated host compartments. Overall we have: $S + \sum I = 1$. Upon primary pathogen exposure, a susceptible host can acquire pathogen type 1 or type 2. The forces of infection (FOI) depend explicitly on prevalence: $\lambda_1(t)$ for type 1, and $\lambda_2(t)$ for type 2. Single carriers block subsequent acquisition of the same type but can acquire the other pathogen type with a reduced rate σ_i, this due to competition between the resident and the newcomer strain. Assuming no clearance, there is an endemic persistence equilibrium whenever $\beta > \mu$ in the absence of intervention ($\rho = 0$). Stable coexistence between types requires further constraints on σ_1 and σ_2, as shown in Fig. S1.

2.3. SIS model for 2 groups of pathogen types (n ≫ 2)

For pathogens with larger antigenic diversity, many types can circulate simultaneously in a host population. The constituent pathogen subtypes can be equivalent in most life-history traits, including basic transmission and clearance potential. However, when vaccination with polyvalent vaccines is considered, it becomes practical to aggregate them into types of *group 1*, (vaccine types, i.e. those that will be targeted by an intervention), and *group 2* (remaining ones, or non-vaccine types). Typically a host may acquire any pathogen type of group 1, or group 2, or be a double carrier of two types: from group 1, from group 2, or one of each.

Acquisition of a second pathogen type generally occurs at a reduced rate compared to single carriage, due to direct competition between the resident and newcomer pathogen types. Competition between any two pathogen types is represented by parameter $\sigma < 1$. When they belong to the same group we apply an extra reduction factor κ to account for the fact that there is one less type available for colonization within the same group. In this sense κ should be seen as a factor representing depletion of available types, which can exert a small or large effect on coinfection by types within the same group, depending on whether the group has many types or just a few.

Although in principle any individual carriage episode can induce some type-specific immunity, we consider the magnitude of such immunity negligible when assessed on the entire pool of *all* pathogen types from the parent group. There is, however, nothing to preclude the inclusion of cumulative immunity in model extensions informed by data on specific pathogens. Meanwhile, we are effectively working with SIS epidemiological dynamics at the level of *groups* of subtypes. The system with vaccination is given by:

$$\begin{cases} \text{Non-vaccinated hosts} \\ \frac{dS^0}{dt} = \mu(1-\rho) - (\lambda_1+\lambda_2+\mu)S^0 + \gamma(1-\rho-S^0) \\ \frac{dI_1^0}{dt} = \lambda_1 S^0 - I_1^0(\sigma\lambda_2 + \kappa\sigma\lambda_1) - (\mu+\gamma)I_1^0 \\ \frac{dI_{2^0}}{dt} = \lambda_2 S^0 - I_2^0(\sigma\lambda_1 + \kappa\sigma\lambda_2) - (\mu+\gamma)I_2^0 \\ \frac{dI_{11}^0}{dt} = \kappa\sigma\lambda_1 I_1^0 - (\mu+\gamma)I_{11}^0 \\ \frac{dI_{22}^0}{dt} = \kappa\sigma\lambda_2 I_2^0 - (\mu+\gamma)I_{22}^0 \\ \frac{dI_{12}^0}{dt} = \sigma(\lambda_2 I_1^0 + \lambda_1 I_2^0) - (\mu+\gamma)I_{12}^0 \end{cases} \quad \begin{cases} \text{Vaccinated hosts} \\ \frac{dS^1}{dt} = \mu\rho - (w\lambda_1+\lambda_2)S^1 - \mu S^1 + \gamma(\rho - S^1) \\ \frac{dI_1^1}{dt} = w\lambda_1 S^1 - I_1^1(\sigma\lambda_2 + w\kappa\sigma\lambda_1) - (\mu+\gamma)I_1^1 \\ \frac{dI_2^1}{dt} = \lambda_2 S^1 - I_2^1(w\sigma\lambda_1 + \kappa\sigma\lambda_2) - (\mu+\gamma)I_2^1 \\ \frac{dI_{11}^1}{dt} = w\kappa\sigma\lambda_1 I_1^1 - (\mu+\gamma)I_{11}^1 \\ \frac{dI_{22}^1}{dt} = \kappa\sigma\lambda_2 I_2^1 - (\mu+\gamma)I_{22}^1 \\ \frac{dI_{12}^1}{dt} = \sigma(\lambda_2 I_1^1 + w\lambda_1 I_2^1) - (\mu+\gamma)I_{12}^1, \end{cases}$$

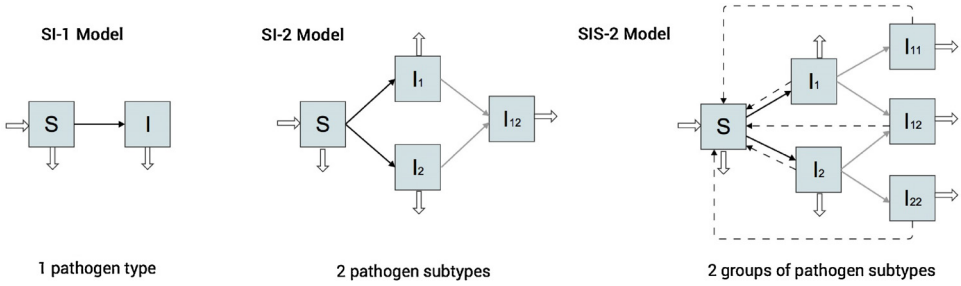


Fig. 1. Model diagrams in the absence of vaccination. The direct competition phenomenon between pathogen subtypes is represented by the grey arrows, modifying the rate of transition from single to dual colonization.

where the parameter γ denotes the clearance rate of each carriage episode, assumed equal for single and dual carriage, as in some pneumococcus models (Colijn et al., 2009; Mitchell et al., 2015). The forces of infection are: $\lambda_1 = \beta(I_1^0 + I_{11}^0 + I_{12}^0/2 + I_1^1 + I_{11}^1 + I_{12}^1/2)$ and $\lambda_2 = \beta(I_2^0 + I_{22}^0 + I_{12}^0/2 + I_2^1 + I_{22}^1 + I_{12}^1/2)$. In this formulation, as well as in the SI-2 formulation, it is assumed that an individual carrying two pathogen subtypes is only 50% infectious for either one of them, a special case of neutral null models (Lipsitch et al., 2009).

Stability of the endemic equilibrium in the absence of intervention requires $R_0 = \beta/(\gamma + \mu) > 1$. A condition for stable coexistence of the two groups of pathogen subtypes in the absence of intervention is that $\kappa < 1$, which is always true given the meaning of κ. As in the previous models, we assume that a fraction ρ of the population is vaccinated, with a vaccine that reduces susceptibility to targeted pathogen types (here group 1) by a factor w ($0 \leq w \leq 1$). The model can be generalized to include finer scales of competition between pathogen types, if important asymmetries happen to be implied by specific pathogen data.

2.4. Inferring vaccine efficacy

In all our transmission models, vaccine efficacy is given by $VE = 1 - w$, varying between 0 and 1, denoting the reduction in probability of pathogen acquisition per contact in vaccinated individuals relative to those non-vaccinated (Haber et al., 1991). We explore these models, with relation to retrospective analyses of cross-sectional prevalence data post-vaccination. Such data may be available through active sampling in observational studies of populations subject to mass-immunization, and we assume they reflect pathogen carriage irrespective of symptoms.

We initially generate cross-sectional prevalence data post-vaccination through numerical simulation of model systems with fixed parameters. We go on to compare the ratio of target pathogen prevalence (type 1 or group 1) in vaccinated vs. non-vaccinated hosts, with the true relative risk w, and we systematically show how the discrepancy between the two varies with time of observation, with transmission intensity and competition parameters between pathogen subtypes. This is performed for the SI-2 and SIS-2 models (but see Supplement Text S2 for an SIR-2 formulation). What we propose as a solution is to fit the full dynamic model with vaccination to prevalence trajectories, and infer in this way vaccine efficacy, simultaneously with other parameters. Notice that this approach is different from the one suggested by Scott et al. (2014), where they advocate that snapshot prevalence ratio itself, or prevalence-odds ratio be used, only at appropriate times post-vaccination. Here we are not proposing a direct use of any static observation but rather a dynamic model fit to multiple temporal observations of relative and absolute prevalence of carriage post-vaccination.

To motivate this approach, we illustrate the discrepancy between prevalence ratio and vaccine efficacy. Thus, we begin by providing analytical insight using the simpler SI-1 model, where it is easy to derive endemic prevalence equilibria pre- and post-vaccination. Subsequently we explore numerically the models with strain competition.

3. Results

3.1. Vaccine efficacy and endemic equilibria of the SI-1 model

Pathogen carriage prevalence at the stable endemic equilibrium in the absence of vaccination (Section 2.1, $\rho = 0$) is given by $I^*_{pre} = 1 - 1/R_0$, where $R_0 = \beta/\mu$. In the presence of vaccination ($\rho > 0$), by setting the differential equations to zero, we can compute the new endemic equilibrium post-vaccine:

$$I^* = \frac{w(R_0 - 1) - 1 + \sqrt{4(1-\rho)R_0(1-w)w + [w(R_0+1)-1]^2}}{2R_0 w}. \quad (1)$$

Prevalence post-vaccine becomes clearly a nonlinear function of basic reproductive number R_0, vaccine efficacy $(1 - w)$, and coverage ρ (Fig. 2a). When the coverage is perfect $\rho = 1$, we have $I^* = 1 - (1/R_0 w)$. One can see for example that to eliminate the pathogen, with perfect coverage, w needs to be higher than $1/R_0$, or with imperfect coverage, the fraction vaccinated ρ and individual vaccine protection w can be jointly traded-off against one another in different critical combinations (i.e. those that satisfy $I^* = 0$, in Eq. (1)).

These analytical expressions also illustrate that if we know the endemic prevalence equilibria pre- and post-vaccine, and the vaccination coverage ρ, we can infer simultaneously R_0 and w, hence vaccine efficacy $(1 - w)$, by comparing pre-vaccine with post-vaccine cross-sectional prevalence of carriage. The above expression can be used to asses the 'overall' effectiveness of a vaccination program, defined as the reduction in the transmission rate for an average individual in a population with a vaccination program at a given level of coverage compared to an average individual in a comparable population with no vaccination program (Halloran et al., 1991; Halloran, 2006).

Similarly, if we are interested in the relative equilibrium prevalence ratio (PR^*) of infection in vaccinated and non-vaccinated hosts, we can also obtain it analytically from the model:

$$PR^* = \frac{I^{1*}/\rho}{I^{0*}/(1-\rho)} = \frac{-1 + w - R_0 w(1-2\rho) + \sqrt{4(1-\rho)R_0(1-w)w + [w(R_0+1)-1]^2}}{2\rho R_0 w}, \quad (2)$$

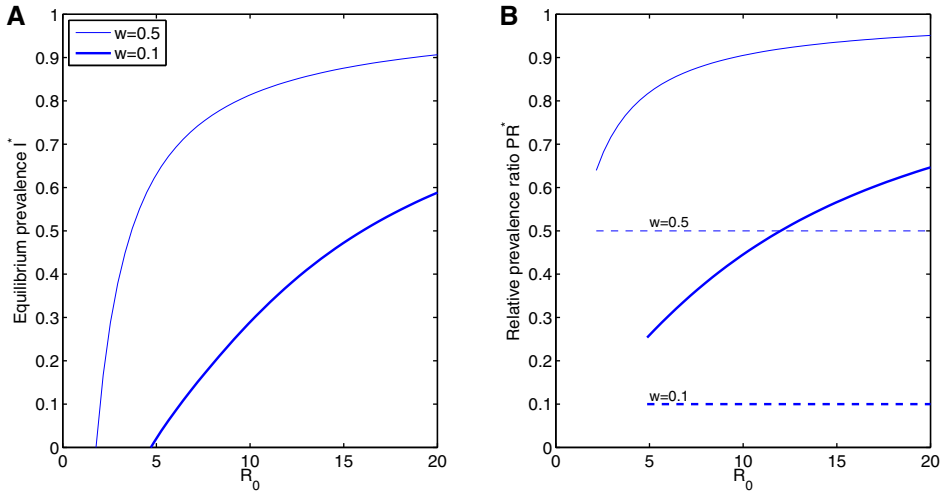


Fig. 2. Infection prevalence at post-vaccination equilibrium vs. transmission intensity (SI-1 model). (A) Absolute prevalence of carriage. (B) Relative prevalence ratio in vaccinated vs. non-vaccinated hosts. The different lines from top to bottom correspond to different values assumed for vaccine efficacy ($VE = 1 - w = 50\%, 90\%$) while keeping fixed coverage $\rho = 0.5$. The ratio of relative infection prevalence in vaccinated vs. non-vaccinated individuals at equilibrium does not reflect the same value of relative risk ($w = 1 - VE$) as transmission intensity R_0 changes. While in this figure, all the points along each line reflect scenarios with the same vaccine efficacy, the prevalence ratio observed is different for each transmission intensity R_0. This shows that one cannot simply translate *1 − prevalence ratio* at equilibrium to a directly interpretable measure of vaccine efficacy, as defined by the $1 - w$ parameter, encapsulating the instantaneous per-unit exposure protection factor experienced by each vaccinated individual. We omit analysis of the corresponding SIS-1 type system, as the results are identical to the SI-1 model.

which reveals that this quantity varies not only with the vaccine parameters ρ and w, but also with transmission intensity, here represented by R_0 (Fig. 2b). This analytical result clearly states that prevalence ratio, even at equilibrium, is unsuitable as a direct indicator of vaccine efficacy, as previously noted (Haber et al., 1991; Shim and Galvani, 2012). In fact, PR^* is commonly greater than w except for special tripartite combinations of (R_0, ρ, w). Nonetheless, it is precisely such nonlinear relationship between R_0, vaccine efficacy, and prevalence, that lies at the heart of typical vaccination programmes against childhood diseases (typically characterized by SIR dynamics), where the critical vaccination coverage needed to eliminate a pathogen has been determined through mathematical models, and applied to control measles, smallpox, mumps, and rubella.

Interestingly, when considering another ratio in our epidemiological model, namely the *prevalence odds ratio* (*POR*), at equilibrium post-vaccine we get:

$$POR^* = \frac{I^{1*}/S^{1*}}{I^{0*}/S^{0*}} = w, \tag{3}$$

confirming the classical result that the prevalence-odds-ratio, at least in a simple SI-1 setting, is a perfect estimator of true relative risk, provided we are at equilibrium. This has been shown previously by studies comparing prevalence ratios and prevalence-odds ratios (Greenland, 1987; Strömberg, 1994; Thompson et al., 1998).

Next, we briefly address how in polymorphic pathogen systems, the relationship between w and prevalence ratio depends systematically also on the strength of competition between pathogen types, and finalize with our proposal to infer w as a fundamental parameter of a dynamic model that can be fitted to cross-sectional data pre- and post-vaccination.

3.2. When prevalence ratio post-vaccination is modulated by strain competition and replacement

Here we consider the effects of subtype competition in pathogen systems with many strains. We simulate hypothetical vaccination scenarios using the dynamic SI-2 and SIS-2 models (Sections 2.2 and 2.3) with two interacting strains or groups of strains for T time units post-vaccination. Time is measured in same units as host life expectancy $1/\mu$. Focusing on the pathogen types targeted by the vaccine, and neglecting the contribution of dual carriers of vaccine and non-vaccine types, the instantaneous prevalence ratio of type 1 or group 1 pathogen, in the two models is defined as:

$$PR(t) = \frac{I_1^1(t)}{I_1^0(t)} \times \frac{1-\rho}{\rho} \quad \text{(SI-2model)}$$

$$PR(t) = \frac{I_1^1(t) + I_{11}^1(t)}{I_1^0(t) + I_{11}^0(t)} \times \frac{1-\rho}{\rho} \quad \text{(SIS-2model)} \tag{4}$$

In the SI-2 model, we consider two possible scenarios: (i) vaccine targets type 1, when type 1 is dominant, and (ii) vaccine targets type 1, when type 2 is dominant prior to intervention. These scenarios are determined by the ratio of competition coefficients $\delta = \sigma_1/\sigma_2$: with type 1 dominance if $\delta < 1$ and type 2 dominance viceversa. When $\sigma_1 = \sigma_2$, the two pathogen types stably coexist at equal abundances. To explore type 1 dominance, in Fig. 3, we set $\sigma_2 = 1$ and consider σ_1 between 0 and 1. To explore type 2 dominance by an equal amount, we set $\sigma_1 = 1$ and vary σ_2.

In agreement with the SI-1 analysis (Fig. 2), the prevalence ratio from the SI-2 model (Eq. (4)) is also most commonly higher than the true relative risk w (Fig. 3). In addition, here we can see that prevalence ratio is very sensitive to the relative magnitudes of competition coefficients, especially when the target strain is dominant. For vaccines that target the dominant type we expect w to be overestimated (vaccine efficacy under-estimated), by a larger amount than for vaccines that target the non-dominant type. This indicates the importance of using flexible model formalisms that accommodate competition parameters to be estimated simultaneously with vaccine efficacy. From the ordering of the curves in this figure, we can expect that using a model that ignores competition ($\sigma_1 = \sigma_2 = 1$) would generally lead to an overestimation of w (underestimation of vaccine efficacy) if type 1 is dominant in the real system, and the opposite if type 2 is dominant.

Intuitively, this can be understood by seeing the vaccine and the intrinsic competitive interactions in the system as two forces that affect the constituent strain dynamics. When type 1 is dominant, and a vaccine targets type 1, the natural tendency of the

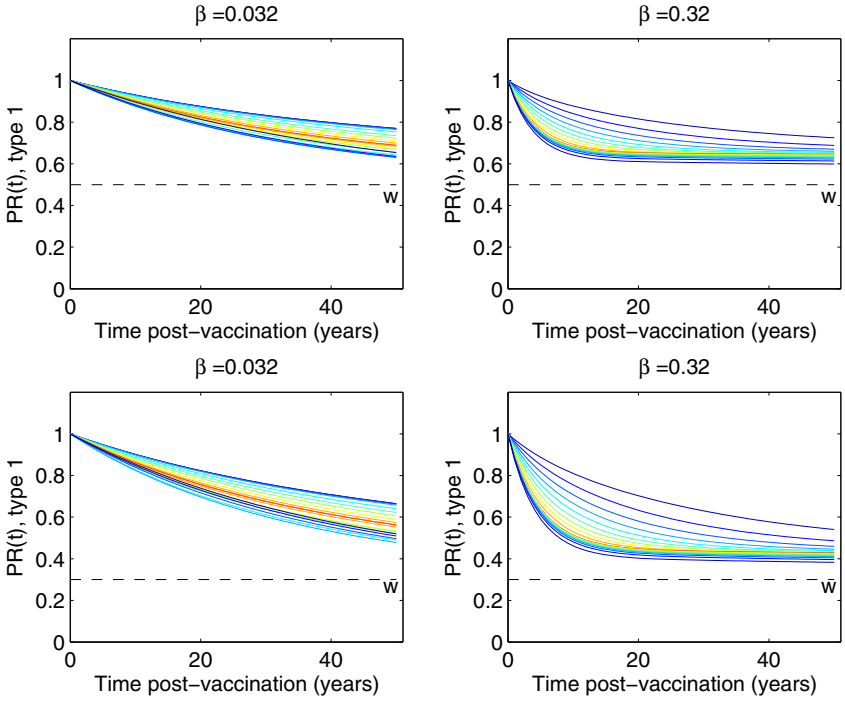


Fig. 3. Prevalence ratio of target type in vaccinated and non-vaccinated hosts vs. true relative risk in the SI-2 model. Type 1 can be dominant prior to vaccination ($0.2 \leq \sigma_1/\sigma_2 \leq 1$), or alternatively, type 2 can be dominant ($0.2 \leq \sigma_2/\sigma_1 \leq 1$). The coloured lines from blue to red correspond to increasing values of the competition ratio, σ_1/σ_2 (above the red curve) and σ_2/σ_1 from 0.2 to 1 (below the red curve). Although as time increases the prevalence ratio tends towards the parameter w, the value remains biased above due to competition between types in both cases, more so if type 1 is the better competitor. Other parameter values: $\mu = 0.0167$, $\rho = 0.5$. Vaccine targets type 1. Different values of w are assumed in top/bottom panels: $w = 0.5$ and $w = 0.3$ respectively. Initial conditions at endemic equilibrium. Time is in units of years, where the average life expectancy is equal to 60 years. The low transmission cases ($\beta = 0.032$), correspond to $R_0 = 1.9$. While the high transmission cases ($\beta = 0.32$) correspond to $R_0 = 19$. For the analogous figure with $PR(t)$ taking into account mixed carriage I_{12} see Fig. S4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

system and the vaccine go in opposite directions, thus the system responds more slowly. When type 2 is dominant instead, the vaccine targeting 1 and direct competition act in the same direction, and we expect faster propagation of vaccine effects in the entire system, whereby also the difference between vaccinated and non-vaccinated hosts emerges earlier ($PR(t) \rightarrow w$ faster post vaccination).

In this model, competition affects also the convergence of *POR* (prevalence odds ratio) of type 1 pathogen to w. We expect from the equilibrium analysis of the SI-1 model, that $POR(t)$ should be closer to the true relative risk w than $PR(t)$. This is what we find. $POR(t)$ tends faster toward w after vaccination is in place in the SI-2 model. However, in the two-strain system with direct competition, $POR(t)$ is also affected by relative strain dominance (Fig. S2), although to a lesser extent than $PR(t)$, and depending on σ_1/σ_2, it also may not always reach asymptotically the true w.

In the SIS-2 model with 2 groups of pathogen types, we uncover a regime where the prevalence ratio of aggregated pathogen types targeted by the vaccine (in single and multiple carriage: $I_1 + I_{11}$) in vaccinated vs. non-vaccinated hosts may be below relative risk, and thus yield an over-estimation of vaccine efficacy if it were to be used for this purpose. This occurs for large transmission intensity β, and κ close to 1 (Fig. 4). When κ is small instead, the pattern $PR(t) > w$ could persist indefinitely. As κ increases, there is initially a downward bias if observations are made too soon after intervention onset, and only after some time does $PR(t)$ tend to the true value of relative risk. Depending on the exact magnitude of β and w, the deviation of relative prevalence ratio from w could persist for a long time after the start of vaccination (Fig. S3). Indeed, as transmission intensity increases, the prevalence of multiple carriage in the host population increases, thus amplifying any indirect effects of competition between types.

Notice that the impact of competition hierarchies on prevalence ratio $PR(t)$ is very sensitive to how this prevalence ratio is defined: whether it takes into account or not, multiple mixed carriage of 1 and 2: I_{12}. In Figs. S4 and S5, we show the analogous scenarios of Figs. 3 and 4, for a prevalence ratio that takes into account the global FOI of target type(s), summing also the contribution $I_{12}/2$ of the mixed carriage host class. The sensitivity of $PR(t)$ to competition hierarchies, in this case decreases in the SI-2 model, and increases drastically in the SIS-2 model, especially for high β, with parallel exacerbated deviation from w, indicating that mixed carriage of target and non-target types contributes more confounding from indirect vaccine effects.

3.3. Using a dynamic model to infer vaccine efficacy

As briefly illustrated above, the use of prevalence ratios to assess vaccine efficacy presents four main problems:

(i) even in the best case of explicitly matching pre-vaccine and post-vaccine *equilibria* (SI-1 model), thus even when satisfying the stationarity assumption, the indirect transmission effects mediated by the vaccine are typically confounded in the prevalence ratio of target types in vaccinated and non-vaccinated individuals (subject to the interplay between R_0 and vaccine coverage);
(ii) in general for observational studies, the prevalence ratio is dependent on the *time of the survey* post-vaccination, which makes it non-robust and hard to compare or interpret across settings, without mechanistically embedding time in the analysis;
(iii) in polymorphic pathogen systems, prevalence-based indicators can be biased depending on the magnitude and direction

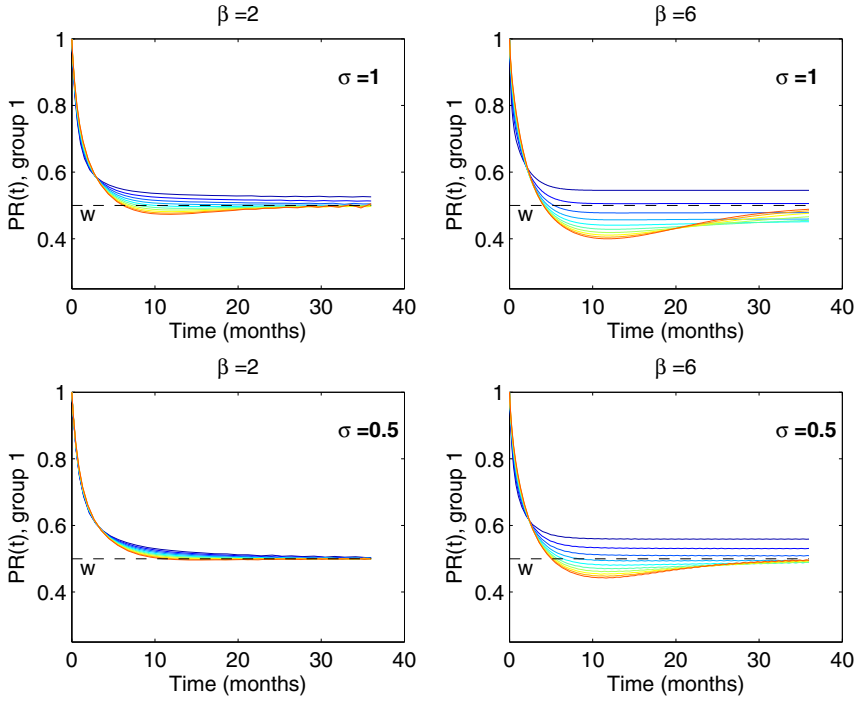


Fig. 4. Deviation of the prevalence ratio from true vaccine-induced protection in the SIS-2 model depends on transmission rate and the magnitude of competition. Higher transmission intensity, leads to larger discrepancy in the immediate time scale after vaccine introduction. Fixed parameter values $\mu = 0.02$, $\rho = 0.5$, $\gamma = 0.6$ and competition coefficient $\sigma = 1$ and 0.5 respectively in the top, and bottom panels. Initial conditions at endemic equilibrium where both groups of types coexist at equal abundance. VE=50% (dashed line depicts $w = 1 - VE$). The depletion factor effecting within-group coinfection κ is varied between 0.1 and 0.9 (coloured lines from blue to red). Increasing overall competition in the system makes indirect effects weaker, thus prevalence ratios reflect more accurately true relative risk w. Time units are months, as here we consider a childhood disease, and the mean age of hosts is 50 months. By type 1 we refer to all pathogen types in group 1 targeted by the polyvalent vaccine. For the analogous figure with $PR(t)$ taking into account mixed carriage I_{12} see Fig. S5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of pre-existing competitive interactions between strains or groups of strains, whose values are hard to know and factor out *a priori*.

(iv) in addition, in polymorphic systems shaped by competition among strains, how the prevalence ratio of target types is defined in terms of *single and multiple carriage of pathogen type combinations* is a critical determinant of the discrepancy with true relative risk, as it is precisely the details of competition at co-colonization that drive indirect effects of the vaccine, especially at high transmission intensities.

Even if we were to use prevalence-odds- ratios (*POR*), instead of prevalence ratios (*PR*), analysis could be inaccurate with regards to estimating true relative risk (and hence vaccine efficacy), as the match between these quantities generally applies only at equilibrium post-vaccine, and it is not exempt from bias introduced by sampling time and competition hierarchies between strains.

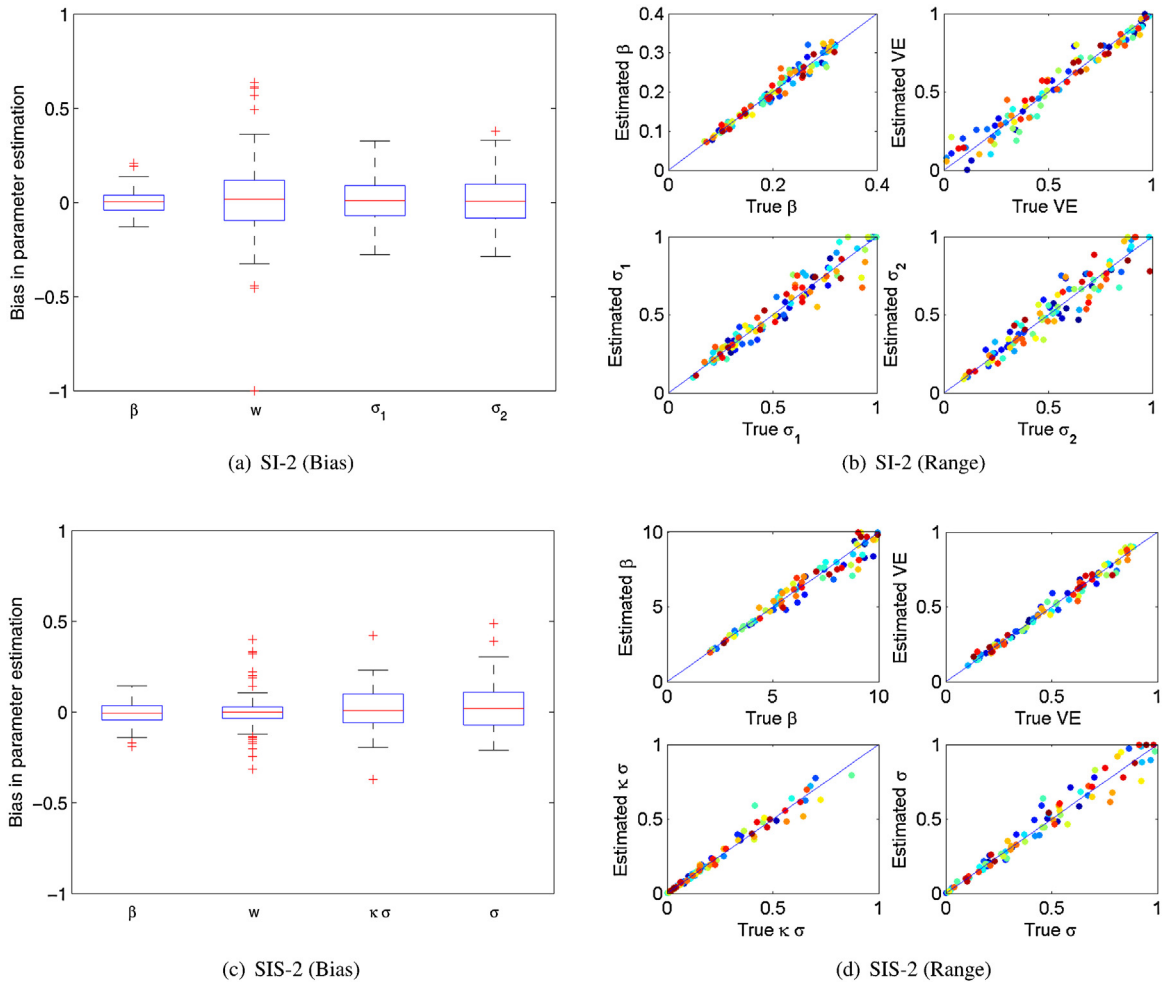
Individual vaccine protection, in the dynamic transmission models ($0 \leq w \leq 1$) multiplies transmission rate (β) and pathogen prevalence, thus it is naturally defined per infectious contact. In contrast, the snapshot prevalence ratio, often reported in field studies, misses the exposure dimension, being an output of the overall dynamics with nonlinear and often non-trivial relation to the original input parameter. To resolve these problems, a productive alternative is to use the full dynamic model, fit it to prevalence data before and after vaccination, as obtained through cross-sectional observational studies, thus estimating together several key epidemiological parameters. Based on pathogen prevalence observations, analysis of epidemiological studies post-licensure could factor out simultaneously the effects of multiple parameters: $\beta, \sigma_1, \sigma_2$ (within the SI-2 model) and $\beta, \kappa\sigma, \sigma$ (in the SIS-2 model), in order to extract the true value of vaccine efficacy ($VE = 1 - w$) against the targeted types. Frequently, epidemiologists will have enough information about the system to define the equations that govern its behaviour, or test simultaneously competing appropriate formulations. In our case, the parameters μ and ρ (and γ in the SIS-2 model) are assumed known.

3.4. Numerical procedure for ODE model fitting and parameter inference

As a proof of concept, we apply the following procedure. We generate hypothetical data performing model simulations with different parameter values, fixing initial conditions at the pre-vaccination endemic state. Observing the state of the system at t_i time units post-vaccination, subsequently we apply nonlinear least-squares optimization (routine *lsqnonlin* in MATLAB) to model-generated trajectories, in order to recover the underlying parameters. Similar methodological approaches exist and are routinely applied to epidemiological data (e.g. in the context of R_0 estimation Cintrón-Arias et al. (2009)). The error function (objective function to be minimized) is given by:

$$Error = \sum_{t_i} \sum_{S,I_1,I_2\ldots} (\text{Theoretical model proportions}(t_i) - \text{Data prevalences}(t_i))^2, \quad (5)$$

where data may be real or synthetically generated, as in our case here. This estimation depends not only on fitting the prevalence of the pathogen types targeted by the vaccine, but also on the prevalence of susceptibles, and of hosts infected with type 2, at specific time points post-vaccination in vaccinated and non-vaccinated



(a) SI-2 (Bias) (b) SI-2 (Range)

(c) SIS-2 (Bias) (d) SIS-2 (Range)

Fig. 5. Bias in parameter estimation using the full epidemiological modelling framework. Dynamics with 100 random parameter combinations were simulated for each model and the nonlinear least-squares optimization was performed on the reduced SI-2 type and SIS-2 system at a specific time post-vaccination. Fixed parameter values were: $\rho = 0.5$, $\mu = 0.0167$ (SI-2) and $\rho = 0.3$, $\mu = 0.02$, $\gamma = 0.57$ (SIS-2). Initial conditions were fixed at endemic coexistence equilibrium in each model. Assuming imperfect observations and a finite population size $N = 1000$, stochasticity was implemented in the observation process by sampling hosts from a multinomial distribution at $t = 0$ and at 3 time-points post-vaccination ($t_1 = 12$, $t_2 = 24$, $t_3 = 36$) which in the SI-2 model reflect 'years' and in the SIS-2 model reflect 'months'. The models were fitted to the resulting 'synthetic' sample proportions. In the bias boxplots ($bias = 1 - (\hat{\theta})/\theta$), the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme bias points not considered outliers, and outliers are plotted individually. β denotes the transmission rate per unit of time, σ_1 and σ_2 denote the direct competition coefficients in the SI-2 model, while σ denotes the competition coefficient and $\sigma\kappa$ the competition coefficient corrected for within-group coinfection in the SIS-2 model.

hosts. The errors are assumed to be normally distributed in this framework, but this assumption can be changed in more sophisticated parameter estimation procedures.

Allowing for imperfect observations (in the synthetic data), realistic sampling error can be added to deterministic simulations. For this, we draw host numbers in different compartments according to a multinomial probability distribution, with probabilities from trajectories of the ODE model (SI-2, SIS-2), and a given sample size N. By applying nonlinear least squares optimization to interpolate such synthetic sample prevalences generated after vaccination, we can recover all the parameters of interest with very good accuracy and in an unbiased manner (Fig. 5).

A critical requirement for the sampling of prevalences post-vaccination is that the sampling time-points are sufficiently spread over the interval $[0, 1/\mu]$, which is intuitive for a vaccine administered at birth, unless its implementation is preceded by a high-coverage vaccination campaign. In order to retrieve sufficient vaccine information, one has to follow the dynamics from the onset of vaccination at least over the entire life-span of a typical individual, i.e. over one generation. Provided this minimal range requirement, the bias decreases further with sample size and with the number of time-points used to sample the population in the post-vaccine era (Fig. 6).

In general, other advantages of deploying dynamic models over more direct statistical descriptions of data, are that they can fill the gaps for unobserved intervals of system behaviour (Fig. 7), generate new hypotheses, and yield more accurate predictions for the future.

3.5. *The role of type-specific immunity and inference in a SIR model*

In the scenarios above, we restricted our attention to infections with no immunity. In a separate model, we extended the basic SI-2 framework, to allow for type-specific immunity upon recovery, all else kept equal, including vaccination against type 1 (see Supplementary Text S2). When changing to an SIR-2 framework, the baseline prevalence of infection in the pre-vaccine population decreases for the same R_0, and competition between types is reduced. We note that prevalence-ratio deviations from true vaccine efficacy persist in the SIR-2 model, and get exacerbated by the oscillatory nature of epidemiological dynamics of infections with immunity (Fig. S6).

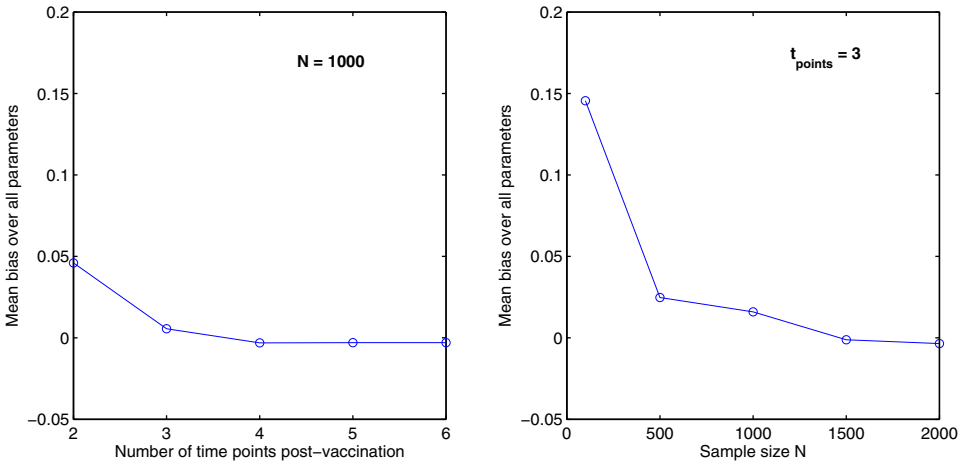


Fig. 6. Bias decreases with number of sampling points and sample size N. Here, we show for the SI-2 model, how the mean (absolute) bias in parameter estimates across different values fitted through the dynamic model, varies with number of sampling time points and sample size. The first time-point is taken at 12 time units post-vaccination, and the subsequent time points are taken in steps of 12. We assume $\mu = 0.0167$, corresponding to a life expectancy of 60 years and for vaccination coverage $\rho = 0.5$. The values of β and σ_1, σ_2 are drawn randomly in the range [0.06, 0.32] and [0, 1], respectively, subject to the constraint of stable endemic coexistence prior to vaccination. As long as the sampling times cover the interval $[0, 1/\mu]$, information about the vaccine which is administered at birth can be extracted.

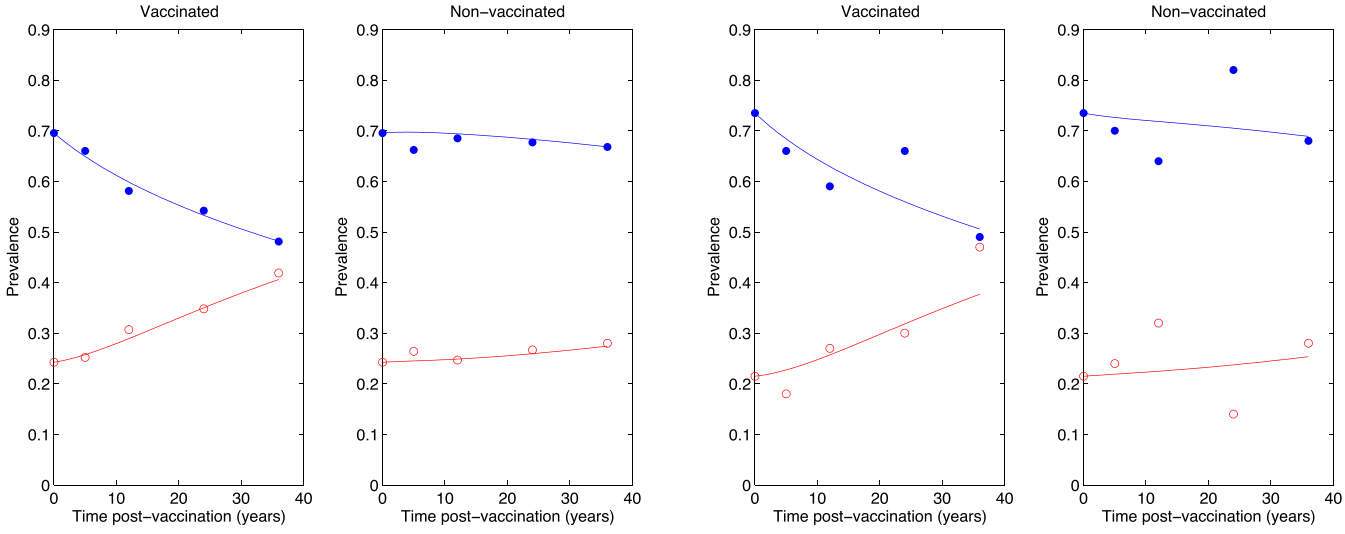


Fig. 7. Illustration of dynamic model fitting to prevalence data in the SI-2 model. After vaccination against type 1 pathogen, cross-sectional prevalence data (circles) can be integrated within a dynamic mathematical model (lines) to estimate epidemiological parameters, such as vaccine efficacy and competition parameters between different pathogen types. Here only overall prevalence of type 1 (blue line, filled circles) and prevalence of type 2 (red line, empty circles) are shown. Parameters used in this simulation: $\beta = 0.3$, $w = 0.2$, $\sigma_1 = 0.1$, $\sigma_2 = 0.5$ and $\rho = 0.5$, $\mu = 0.0167$ are fixed. $N = 1000$ in (a) and $N = 100$ in (b) has been used in the multinomial sampling scheme to add sampling error to synthetic data. Applying the nonlinear least squares routine, in these particular instances, we have obtained these point estimates: (a) $\beta = 0.297$, $w = 0.215$, $\sigma_1 = 0.108$, $\sigma_2 = 0.441$ and (b) $\beta = 0.416$, $w = 0.184$, $\sigma_1 = 0.059$, $\sigma_2 = 0.283$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

When applying dynamic model fitting, we were able to infer the parameters of interest $\beta, w, \sigma_1, \sigma_2$, also for the SIR-2 model from prevalence observations post-vaccination. We grouped 'data' into: pathogen-free hosts (susceptible and immune), those carrying type 1, type 2, and those carrying both 1 and 2. Under random sampling effects (using the multinomial framework), even though transmission rate and vaccine efficacy could be inferred accurately, the quantification of the direct competition coefficients σ_1 and σ_2 was harder (Fig. S7). This is unsurprising, given the lower pathogen prevalence expected in the SIR-2 model, as a large proportion of hosts eventually become immune to both types and competition information is lost. In fact, we expect that estimation of direct competition parameters in a SIR framework requires larger sample sizes and more time-points, or a complete resolution of ODE variables with regards to host immune status and serological history.

Taken together, our simulations convey that in principle, dynamic model fitting through interpolating pre-vaccine and post-vaccine pathogen prevalences can be used to retrospectively assess vaccine effects in different host populations, under different competition scenarios between target and non-target strains, accounting for essential nonlinearities in transmission.

4. Discussion

In this study, we have suggested to close the gap between mechanistic models of vaccine effects and statistical approaches for estimation of vaccine efficacy, addressing in addition multi-type pathogens characterized by direct competition. One reason for why mechanistic models have been somewhat neglected in vaccine assessment studies, is that for validation, dynamic

epidemiological models have to rely on spatiotemporally resolved data, and to date, epidemiological data tend to be static, and therefore more amenable to the tools of statistics. Yet, with increasing infrastructures and efforts to monitor health in populations over time, comes an opportunity to apply more dynamic frameworks in the future, and match them with dynamic vaccine study designs.

4.1. The importance and challenges of dynamic modelling

Here, we draw attention on a new application of mathematical models: namely, in the inference of vaccine efficacy from cross-sectional prevalence data, accounting for other critical determinants of pathogen dynamics such as transmission rate, vaccine coverage, and competition between multiple types (Choi et al., 2011; Gjini et al., 2016). These data can be gathered from active sampling and cross-sectional population surveys pre- and post-vaccination, national screening programs after mass immunization, or reported pathogen prevalence in different countries under specific vaccination coverage rates.

We focused on scenarios of treatment effect homogeneity (leaky vaccine) and baseline homogeneity (no structure in the host population), assuming randomization across vaccinated and non-vaccinated hosts. Of course the models can be shaped so as to incorporate alternative scenarios, but this was beyond our scope here. Depending on the type of data that are available (e.g. if prevalence information is stratified by age, geography etc.), epidemiological models can include transmission between age groups or different spatial locations (e.g. as in (Goeyvaerts et al., 2015)). The spirit of inference remains the same; only in those cases one would have to specify contact parameters and feedbacks between different sub-populations. In the current paper, the main structuring dimension was host vaccination status.

The need to modify current vaccine evaluation practices by including the exposure dimension in vaccine studies has been recently pointed out by Gomes et al. (2014), and additional computational frameworks are already emerging to estimate per-exposure vaccine effects at the individual level (O'Hagan et al., 2014). Interpretation of vaccine efficacy is inevitably entangled with context-specific epidemiological interactions and feedbacks. Thus, accurate quantification of such individual protection parameter, which is robust, identifiable and comparable across settings, can benefit from inclusion of more mathematical approaches in vaccine studies. Any length of post-vaccine follow-up period can be interpreted within dynamic models, as long as the time-scale of observations covers the typical host life-span interval $[0, 1/\mu]$, especially in the case of *SI* and *SIR* dynamics, and as long as approximations underlying model structure (e.g. homogeneous mixing) are valid.

Notice, that the shortcoming of having to sample over an interval of $1/\mu$ time-units to extract epidemiological parameters and vaccine efficacy, if necessary, can be dealt with through parameter inference procedures that extract baseline parameters such as transmission rate and competition coefficients from fitting analytical equilibrium expressions to prevalences pre-vaccination, and then using those estimates to project dynamics forward in the vaccine era. In that case, with already known coverage rate and known baseline parameters, inference of vaccine efficacy alone through dynamic fitting of prevalence observations over shorter time-scales should prove easier.

Depending on the transmission model, care must be taken with regards to correlation between parameters. Some epidemiological quantities, e.g. total prevalence of target types, may display the same global magnitude at a given time post-vaccination, for very different parameter combinations (e.g. κ and *VE* in the SIS-2 model, shown in Fig. S8). However, if finer-scale data are available, including stratification with respect to host vaccination status and single/multiple carriage, the apparently correlated parameters can then be separated. This highlights the importance of high resolution and high quality epidemiological data for parameter estimation.

Here we show that transmission rate and vaccine efficacy against targeted pathogen types can be robustly estimated by dynamic model fitting, even in the presence of observation error. Yet, we also show that our power to identify direct competition coefficients from post-vaccine prevalence data may depend on the overall level of pathogen prevalence in the population. Infectious agents that display lower prevalence are likely to require larger sample sizes in prevalence studies, for a correct resolution of their diversity and inference of underlying subtype competition. This is especially relevant in systems where hosts recover with immunity. Lack of estimability of interaction coefficients however does not necessarily mean the model assuming those interactions is invalid; rather it may hint that those particular parameters are not as important for the overall dynamics at the population scale considered (Wikramaratna et al., 2014), or that a better resolution of host immunological status, besides simply carriage status, is needed.

4.2. Vaccine efficacy and effectiveness: linking trials and population observations

Notice that while efficacy is a direct parameter, that is easily understood and can be averaged across settings, vaccine effectiveness is a somewhat more subjective criterion: (i) it can be defined over a specific time post-vaccination, e.g. reduction in overall prevalence in the 3 years following vaccination; (ii) in the case of multi-type systems, effectiveness can be defined with regards to carriage of the targeted pathogen types, or total carriage; (iii) it can be defined with regards to the vaccinated sub-population only, or as an overall measure for the entire population irrespective of vaccination status; (iv) it can reflect disease manifestations of the pathogen at the population level or asymptomatic carriage states. All these 'effectiveness' indicators are important outputs of the epidemiological dynamics, that can be calculated to meet the demands of various policy-makers, and while they may be different, the intrinsic vaccine protection *VE* per individual is the same.

When linking vaccine efficacy against acquisition obtained through randomized controlled trials pre-licensure, and the vaccine efficacy estimated dynamically post-licensure, we expect in principle the two quantities should match. That is why the framework we propose could also serve as a validation test. Notice however that often vaccine trials are conducted in one population, while vaccines are implemented in another. Thus, differences could arise. For example, interference with maternal antibodies against local pathogens may alter the vaccine protection against other pathogens in a new vaccination programme. Immunological responses may also be influenced by nutrition status and other factors, thus there is nothing to preclude post-licensure assessment of *VE* in populations where trials were not conducted in the first place. All these issues can be explored using dynamic models of transmission to interpret vaccination-induced changes in populations, and extract the basic protection parameter. Even if the two quantities do not match, one could argue that studying their discrepancy would open the way for better data collection, model improvement, fine-tuning of vaccination coverage, inclusion of host heterogeneities, and ultimately a better understanding of the dynamics. Ideally, one should work together with the two approaches: both pre-licensure and post-licensure, and try to match predictions with retrospective analyses. Although not primarily intended to address waning of immunity, a dynamic framework is ideal to explore also the issue of

the waning of vaccine-mediated protection, especially when interpolating prevalence observations over long intervals since the onset of a vaccination programme.

A dynamic model, by its nature of mechanistically connecting multi-dimensional epidemiological observations over time, can explain and link also more kinds of data, including the prevalence of multiple carriage, of competing pathogen types, and alterations in prevalence of total carriage over time. All these quantities could have implications for disease, for the evolutionary potential of the pathogen, and interactions with other pathogens. A dynamic model thus offers a broader view of 'effectiveness'. Predictions for pathogen dynamics in the population can be further integrated into mechanistic models for specific symptoms, as shown by Rodriguez-Barraquer et al. (2013).

4.3. Methodological prospects

Our aims in this paper were illustrative and conceptual. For this reason, we did not elaborate extensively on methodological aspects. Applying dynamic model fitting within a Bayesian framework, to account for parameter uncertainty, is also possible. One could use the same objective function as in Section 3.4, or an explicit multinomial likelihood function, for the exact numbers of hosts observed in different epidemiological classes. In the latter case, the sample size N and expected probability vector coming from numerical integration of the ODE system with its set of parameters, would enter explicitly the likelihood term. Moreover, prior information could be added about specific parameters, if available. One such Bayesian procedure is explained in detail, and applied to a specific dataset on pneumococcus vaccination in Portugal, in a related paper by the authors (Gjini et al., 2016), where in addition an analysis of competing model formulations is undertaken. Another important factor is process noise and stochasticity (Ellner et al., 1998; Shrestha et al., 2011). Parameter inference in mechanistic models based on systems of coupled ODEs is a timely and computationally challenging problem. Many advances in this respect are being made across fields such as statistics and computing (Haario et al., 2006; Calderhead et al., 2009; Girolami and Calderhead, 2011; Dondelinger et al., 2013; King et al., 2014), and applied successfully across domains of science, including systems biology, astrophysics and climate studies. Refinement of such methodologies for epidemiological models with vaccination should be straightforward given the multitude of tools that already exist or are in development. The conceptual backbone of dynamic inference is likely to apply across systems, however, challenges in specific diseases will inevitably involve adaptation of model structure and refinement of the inference procedure.

Finally, there are many ways in which two pathogen strains can compete with each other, most notably via cross-immunity, which we did not consider here. The existence of natural cross-immunity, such as among dengue (Adams et al., 2006; Wearing and Rohani, 2006), or Human Papilloma Virus strains (Elbasha and Galvani, 2005; Durham et al., 2012), poses the issue of eventual cross-immunity also in the vaccine-induced protection against particular pathogen types. To account for immune-mediated interactions, models need to specify how naturally acquired type-specific immunity might interfere with vaccine-induced immunity (Gomes et al., 2004) and vaccine effectiveness (Omori et al., 2012). Robust inference of vaccine protection in these cases remains an open area for future research, calling for sophisticated vaccine study designs, as do other factors, such as host population structure, secular trends and stochasticity. An accurate understanding of intervention effectiveness in populations under mass immunization will increasingly require integrated modelling frameworks for pre- and post-vaccine dynamics, accounting for the interplay of all these factors.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.epidem.2015.11.001.

References

Adams, B., Holmes, E., Zhang, C., Mammen, M., Nimmannitya, S., Kalayanarooj, S., Boots, M., 2006. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. Proc. Natl. Acad. Sci. U.S.A. 103 (38), 14234–14239.

Agnandji, S.T., Lell, B., Soulanoudjingar, S.S., Fernandes, J.F., Abossolo, B.P., Conzelmann, C., Methogo, B., Doucka, Y., Flamen, A., Mordmüller, B., et al., 2011. First results of phase 3 trial of RTS, S/AS01 malaria vaccine in African children. N. Engl. J. Med. 365 (20), 1863–1875.

Anderson, R.M., May, R.M., Anderson, B., 1992. Infectious Diseases of Humans: Dynamics and Control, vol. 28. Wiley Online Library.

Andrews, N., Waight, P.a., Borrow, R., Ladhani, S., George, R.C., Slack, M.P.E., Miller, E., 2011. Using the indirect cohort design to estimate the effectiveness of the seven valent pneumococcal conjugate vaccine in England and Wales. PLoS ONE 6 (12), e28435.

Andrews, N.J., Waight, P.A., Burbidge, P., Pearce, E., Roalfe, L., Zancolli, M., Slack, M., Ladhani, S.N., Miller, E., Goldblatt, D., 2014. Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: a postlicensure indirect cohort study. Lancet Infect. Dis. 14 (9), 839–846.

Biondi, B.E., Weiss, S.H., 2015. Impact of bivalent human papillomavirus (HPV) vaccination upon the risk of acquisition of other HPV types. Med. Res. Arch. 1 (1).

Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J.R., Elvin, L., Ensor, K.M., Hackell, J., Siber, G., et al., 2000. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Pediatric infect. Dis. J. 19 (3), 187–195.

Calderhead, B., Girolami, M., Lawrence, N.D., 2009. Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. In: Advances in Neural Information Processing Systems, pp. 217–224.

Choi, Y.H., Jit, M., Gay, N., Andrews, N., Waight, P.A., Melegaro, A., George, R., Miller, E., 2011. 7-Valent pneumococcal conjugate vaccination in England and Wales: is it still beneficial despite high levels of serotype replacement. PLoS ONE 6 (10), e26190.

Cintrón-Arias, A., Castillo-Chávez, C., Bettencourt, L.M., Lloyd, A.L., Banks, H., 2009. The estimation of the effective reproductive number from disease outbreak data. Math. Biosci. Eng. 6 (2), 261–282.

Colijn, C., Cohen, T., Fraser, C., Hanage, W., Goldstein, E., Givon-Lavi, N., Dagan, R., Lipsitch, M., 2009. What is the mechanism for persistent coexistence of drug-susceptible and drug-resistant strains of streptococcus pneumoniae? J. R. Soc. Interface, rsif20090400.

Comanducci, M., Bambini, S., Brunelli, B., Adu-Bobie, J., Aricò, B., Capecchi, B., Giuliani, M.M., Masignani, V., Santini, L., Savino, S., et al., 2002. NadA, a novel vaccine candidate of *Neisseria meningitidis*. J. Exp. Med. 195 (11), 1445–1454.

Crowe, E., Pandeya, N., Brotherton, J.M., Dobson, A.J., Kisely, S., Lambert, S.B., Whiteman, D.C., et al., 2014. Effectiveness of quadrivalent human papillomavirus vaccine for the prevention of cervical abnormalities: case–control study nested within a population based screening programme in Australia. BMJ 348, g1458.

del Angel, R.M., Reyes-del Valle, J., 2013. Dengue vaccines: strongly sought but not a reality just yet. PLoS Pathog. 9 (10), e1003551.

Dondelinger, F., Filippone, M., Rogers, S., Husmeier, D., 2013. ODE Parameter Inference Using Adaptive Gradient Matching with Gaussian Processes.

Durham, D.P., Poolman, E.M., Ibuka, Y., Townsend, J.P., Galvani, A.P., 2012. Reevaluation of epidemiological data demonstrates that it is consistent with cross-immunity among human papillomavirus types. J. Infect. Dis. 206 (8), 1291–1298.

Elbasha, E.H., Galvani, A.P., 2005. Vaccination against multiple HPV types. Math. Biosci. 197 (1), 88–117.

Ellner, S., Bailey, B., Bobashev, G., Gallant, A., Grenfell, B., Nychka, D., 1998. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. Am. Nat. 151 (5), 425–440.

Girolami, M., Calderhead, B., 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. J. R. Stat. Soc.: Ser. B Stat. Methodol. 73 (2), 123–214.

Gjini, E., Valente, C., Sa-Leao, R., Gomes, M., 2016. How direct competition shapes coexistence and vaccine effects in multi-strain pathogen systems. J. Theoret. Biol. 388, 50–60.

Goeyvaerts, N., Willem, L., Van Kerckhove, K., Vandendijck, Y., Hanquet, G., Beutels, P., Hens, N., 2015. Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season-specific influenza-like illness incidence. Epidemics 13, 1–9.

Gomes, M.G.M., Lipsitch, M., Wargo, A.R., Kurath, G., Rebelo, C., Medley, G.F., Coutinho, A., 2014. A missing dimension in measures of vaccination impacts. PLoS Pathog. 10 (3).

Gomes, M.G.M., White, L.J., Medley, G.F., 2004. Infection, reinfection, and vaccination under suboptimal immune protection: epidemiological perspectives. J. Theoret. Biol. 228 (4), 539–549.

Greenland, S., 1987. Interpretation and choice of effect measures in epidemiologic analyses. American journal of epidemiology 125 (5), 761–768.

Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. Stat. Comput. 16 (4), 339–354.
Haber, M., Longini, I.M., Halloran, M.E., 1991. Measures of the effects of vaccination in a randomly mixing population. Int. J. Epidemiol. 20 (1), 300–310.
Halloran, M., 2006. Overview of vaccine field studies: types of effects and designs. J. Biopharmaceut. Stat. 16 (4), 415–427.
Halloran, M.E., Haber, M., Longini, I.M., Struchiner, C.J., 1991. Direct and indirect effects in vaccine efficacy and effectiveness. Am. J. Epidemiol. 133 (4), 323–331.
Halloran, M.E., Longini, I.M., Struchiner, C.J., Longini, I.M., Struchiner, C.J., 2010. Design and Analysis of Vaccine Studies. Springer.
Heesterbeek, J., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, vol. 5. Wiley.
Hernán, M.A., 2010. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.) 21 (1), 13.
Insinga, R.P., Dasbach, E.J., Elbasha, E.H., Puig, A., Reynales-Shigematsu, L.M., 2007. Cost-effectiveness of quadrivalent human papillomavirus (HPV) vaccination in Mexico: a transmission dynamic model-based evaluation. Vaccine 26 (1), 128–139.
Keeling, M.J., 2005. Models of foot-and-mouth disease. Proc. R. Soc. Lond. B: Biol. Sci. 272 (1569), 1195–1202.
King, A.A., Nguyen, D., Ionides, E.L., 2014. Statistical inference for partially observed Markov processes via the r package pomp. J. Stat. Softw. 55 (2).
Lipsitch, M., 1997. Vaccination against colonizing bacteria with multiple serotypes. Proc. Natl. Acad. Sci. U.S.A. 94 (12), 6571–6576.
Lipsitch, M., Colijn, C., Cohen, T., Hanage, W.P., Fraser, C., 2009. No coexistence for free: neutral null models for multistrain pathogens. Epidemics 1 (1), 2–13.
Martcheva, M., Bolker, B.M., Holt, R.D., 2008. Vaccine-induced pathogen strain replacement: what are the mechanisms? J. R. Soc. Interface 5 (18), 3–13.
Mitchell, P.K., Lipsitch, M., Hanage, W.P., 2015. Carriage burden, multiple colonization and antibiotic pressure promote emergence of resistant vaccine escape pneumococci. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 370 (1670), 20140342.
Moberley, S., Andrews, R., 2014. An evaluation of the indirect cohort method to estimate the effectiveness of the pneumococcal polysaccharide vaccine. J. Vaccines Immunol. 2, 4–6.
O'Hagan, J.J., Lipsitch, M., Hernán, M.A., 2014. Estimating the per-exposure effect of infectious disease interventions. Epidemiology 25 (1), 134–138.
Omori, R., Cowling, B.J., Nishiura, H., 2012. How is vaccine effectiveness scaled by the transmission dynamics of interacting pathogen strains with cross-protective immunity? PLoS ONE 7 (11), e50751.
Richardson, D., 2004. An incidence density sampling program for nested case-control analyses. Occup. Environ. Med. 61 (12), e59.
Rinta-Kokko, H., Dagan, R., Givon-Lavi, N., Auranen, A., 2009. Estimation of vaccine efficacy against acquisition of pneumococcal carriage. Vaccine 27 (29), 3831–3837.
Rodriguez-Barraquer, I., Mier-y Teran-Romero, L., Burke, D.S., Cummings, D.A., 2013. Challenges in the interpretation of dengue vaccine trial results. PLoS Neglect. Trop. Dis. 7 (8), e2126.
Sabchareon, A., Wallace, D., Sirivichayakul, C., Limkittikul, K., Chanthavanich, P., Suvannadabba, S., Jiwariyavej, V., Dulyachai, W., Pengsaa, K., Wartel, T.A., et al., 2012. Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. Lancet 380 (9853), 1559–1567.
Scott, P., Herzog, S.A., Auranen, K., Dagan, R., Low, N., Egger, M., Heijne, J.C., 2014. Timing of bacterial carriage sampling in vaccine trials: a modelling study. Epidemics 9, 8–17.
Shim, E., Galvani, A.P., 2012. Distinguishing vaccine efficacy and effectiveness. Vaccine 30 (47), 6700–6705.
Shrestha, S., King, A.A., Rohani, P., 2011. Statistical inference for multi-pathogen systems. PLoS Comput. Biol. 7 (8), e1002135.
Strömberg, U., 1994. Prevalence odds ratio vs prevalence ratio. Occup. Environ. Med. 51 (2), 143.
Thompson, M.L., Myers, J., Kriebel, D., 1998. Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done? Occup. Environ. Med. 55 (4), 272–277.
van Boven, M., Ruijs, W.L., Wallinga, J., O'Neill, P.D., Hahne, S., 2013. Estimation of Vaccine Efficacy and Critical Vaccination Coverage in Partially Observed Outbreaks.
Wearing, H.J., Rohani, P., 2006. Ecological and immunological determinants of dengue epidemics. Proc. Natl. Acad. Sci. U.S.A. 103 (31), 11802–11807.
Weinberger, D.M., Malley, R., Lipsitch, M., 2011. Serotype replacement in disease after pneumococcal vaccination. Lancet 378 (9807), 1962–1973.
Wikramaratna, P.S., Kucharski, A., Gupta, S., Andreasen, V., McLean, A.R., Gog, J.R., 2014. Five challenges in modelling interacting strain dynamics. Epidemics.